CLAIMS

What is claimed is:

1. A computerized method of identifying a statistically significant group of two or more genomic datums in the form of alleles and/or SNP patterns as these genomic datums affect given clinical results, which group is generally known as a clinically relevant alleles combination and/or characteristic SNP pattern as the case may be, the method comprising:

obtaining numerous examples of (i) clinical alleles and/or SNP pattern genomic data, and (ii) historical clinical results corresponding to this genomic data;

constructing a neural network suitable to map (i) the allele and/or SNP pattern genomic data as inputs to the neural network to (ii) the historical clinical results as outputs of the neural network;

exercising the constructed neural network to so map (i) the clinical alleles and/or SNP pattern genomic data as inputs to (ii) the historical clinical results as outputs; and

conducting an automated procedure to vary the mapping function, inputs to outputs, of the constructed and exercised neural network in order that, by minimizing an error measure of the mapping function, a more optimal neural network mapping architecture is realized;

wherein realization of the more optimal neural network mapping architecture means that any irrelevant inputs are effectively excised, meaning that the more optimally mapping neural network will substantially ignore input alleles and/or SNP pattern genomic data that is irrelevant to output clinical results; and

wherein realization of the more optimal neural network mapping architecture also means that any relevant inputs are effectively identified, making that the more optimally mapping neural network will serve to identify, and use, those input alleles and/or SNP pattern genomic data that are relevant, in combination, to output clinical results.

2. The computerized method of identifying a clinically relevant combination of genomic datums in the form or alleles and/or SNP patterns according to claim 1 wherein the conducting of an automated procedure to vary the neural network mapping function comprises:

varying the architecture of the neural network by a genetic mapping algorithm.

3. The computerized method of identifying a clinically relevant combination of genomic datums in the form or alleles and/or SNP patterns according to claim 1 wherein the obtaining is of numerous examples of (i) alleles datums of types taken from a first group consisting essentially of:

entire gene families;
specific alleles;
specific base pair sequences;
locations and types of introns; and
nucleotide polymorphism,

plus at least three members of a second, environmental, group consisting essentially of:

diet type;
home region;
occupation;
viral levels;
peptide levels;
blood plasma levels;
pharmacokinetic and pharmacodynamic parameters.

4. The computerized method of identifying a clinically relevant combination of genomic datums in the form or alleles and/or SNP patterns according to claim 3 wherein the obtaining of numerous examples of (i) alleles data is of alleles data further including genetic data regarding

ethnicity.

5. The computerized method of identifying a clinically relevant combination of genomic datums in the form or alleles and/or SNP patterns according to claim 1 wherein the obtaining of numerous examples of (i) alleles data is of data from a first group consisting essentially of:

entire gene families,
specific alleles,
specific base pair sequences,
locations and types of introns, and
nucleotide polymorphism;

plus at least two members of an at-least-partially-environmentally-determined second group consisting essentially of:

diet type,
home region,
occupation,
viral levels,
peptide levels,
blood plasma levels, and
pharmacokinetic and pharmacodynamic parameters;

plus at least one member of a third group, which third group members are determined by a combination of genetic and environmental factors, consisting essentially of

ethnicity, and
race.

6. The computerized method of identifying a clinically relevant combination of genomic datums in the form or alleles and/or SNP patterns according to claim 1 wherein the obtaining of numerous examples of (ii) clinical results data is of clinical results data from a group consisting essentially of:

presence of any of biological conditions, diseases and characteristics;

quantitative clinical measures of a patient;

any presence of characteristics for which a genetic or environmental origin is, as of January 1, 2000, either not clear or not uniquely defined, including aggressive tendencies, sexual orientation, and eating disorders, all of which characteristics are called sociological variables; and

cost or performance functions calculated from values of multiple "real" clinical variables.

7. A method of identifying a clinically relevant alleles combination comprising:

1) obtaining a set of examples of (i) alleles data from the group consisting essentially of genomic data from the group consisting essentially of

entire gene families,
specific alleles,
specific base pair sequences,
locations and types of introns, and
nucleotide polymorphism,

plus at least one member of an at-least-partially-environmentally-determined group consisting essentially of

diet type,
home region,
occupation,
viral levels,
peptide levels,
blood plasma levels, and
pharmacokinetic and pharmacodynamic parameters,

plus at least one member of a group determined by a combination of genetic and environmental factors consisting essentially of

ethnicity,

plus corresponding (ii) clinical results data from the group consisting essentially of

presence of any of biological conditions, diseases and characteristics,

quantitative clinical measures of a patient,

any presence of characteristics for which a genetic or environmental origin is, as of January 1, 2000, either not clear or not uniquely defined, including aggressive tendencies, sexual orientation, and eating disorders, which characteristics are called sociological variables, and

cost or performance functions calculated from values of multiple "real" clinical variables;

2) constructing a neural network to map the (i) alleles data as inputs to the (ii) clinical results data as outputs; and

3) training by and with an automated neural network training program the constructed neural network so as to optimize a measure of fitness, being an error measure of the neural network, the training permitting variation in an architecture of the constructed neural network, said neural network architecture including at least numbers and identities of inputs actually fed to the neural network;

wherein variation of at least the numbers and identities of inputs, being (i) alleles data, that is actually fed to the neural network so as to optimally correlate to output data, being (ii) clinical results data, so as to optimize the measure of fitness makes that the trained neural network is fit to relate input (i) alleles data to output (ii) clinical data, and does thus show which of the alleles inputs are essentially irrelevant as insignificantly affect clinical results, and which of the alleles inputs are, in combination, significant to clinical results;

wherein training of the neural network serves to identify clinically relevant alleles combinations.

8. The method of identifying a clinically relevant alleles combination according to claim 7 wherein the training of the constructed neural network is by and with an automated neural network training program comprising:

a programmed genetic algorithm.

9. A method of identifying from the genomic data of an individual organism an adverse reaction to a therapy for at least one disease of the organism,

the method particularly serving to identify a relationship between, on the one hand, (i) any adverse reaction to at least one therapy for at least one disease of an organism, and, on the other hand, genomic data of the organism in the form of two or more alleles and/or SNP pattern(s) of the organism,

the method still more particularly serving to determine which of a large number of alleles as variously occur in the genomic data of a large number of individual organisms are, in actual fact, relevant, both individually and in combination, to certain biological and social variables of these organisms, including the adverse reaction to the at least one therapy for the at least one disease of these organisms,

the method comprising:

1) constructing a neural network suitable to map (i) genomic data of individual organisms as inputs to (ii) historical incidences of responses, including adverse reactions, to therapies for diseases of the individual organisms as outputs;

2) training the constructed neural network on numerous examples of (i) genomic data, as corresponds to (ii) historical incidences of responses including adverse reactions to therapies for the diseases of a multiplicity of individual organisms, so as to make a trained neural network that is fit, and that possesses a measure of goodness, to map (i) genomic data to (ii) incidences of therapeutic responses, including adverse reactions, to therapies for the diseases of the organisms; and

3) exercising the trained constructed neural network in respect of a particular therapy for a particular disease of a particular organism, from among the therapies and the diseases to which the neural network was trained for organism including the particular organism, in order to identify any relationship between (i) any adverse reaction among the responses to the particular therapy, and (ii) genomic makeup of the particular organism;

wherein the neural network is constructed for, and trained on, more organisms than the individual organism on which it is exercised.

10. A method of predicting an optimal drug dosage and/or drug efficacy for a particular individual patient in respect of genomic data, including alleles and/or characteristic SNP patterns, of the particular individual patient, the method comprising:

training a neural network on numerous examples of (i) genomic data including alleles and/or characteristic SNP patterns, and corresponding (ii) historical drug dosage results including optimal drug dosages, for a multiplicity of patients so as to make a trained neural network that is fit, and that possesses a measure of goodness, to map (i) genomic data, including alleles and/or characteristic SNP patterns, to (ii) drug dosage results including optimal drug dosages; and

exercising the trained neural network on the genomic data, including the alleles and/or characteristic SNP patterns, of a particular individual patient to predict an optimal drug dosage for the particular individual patient from among the optimal drug dosages to which the neural network was trained.

11. A method of identifying from the genomic data of an individual organism a suitable therapy for at least one disease of the individual organism,

the method particularly serving to identify a relationship between, on the one hand, at least one therapy for at least one disease of an organism, and, on the other hand, genomic data of the organism in the form of two or more alleles and/or SNP pattern(s) of the organism,

the method still more particularly serving to determine which of a large number of alleles as variously occur in the genomic data of a large number of individual organisms are, in actual fact, relevant, both individually and in combination, to certain biological and social variables of these organisms, including the efficacy of at least one therapy to at least one disease of these organisms,

the method comprising:

1) constructing a neural network suitable to map (i) genomic

data in the form or two or more alleles and/or SNP patterns of individual organisms as inputs to (ii) historical incidences of responses to therapies for diseases of the individual organisms as outputs; and

2) training the constructed neural network on numerous examples of (i) genomic data as corresponds to (ii) historical incidences of responses to therapies for the diseases of a multiplicity of individual organisms so as to make a trained neural network that is fit, and that possesses a measure of goodness, to map (i) said genomic data to (ii) said incidences of responses to therapies for the diseases of the organisms; and

3) exercising the trained constructed neural network in respect of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the neural network was trained, in order to identify a relationship between the particular therapy and genomic data, in the form of two or more alleles, of the organisms.

12. A method of identifying and predicting from the genomic data of an individual organism susceptibility of the organism to a disease,

the method more particularly serving to identify and predict susceptibility of a particular individual patient to at least one disease in respect of alleles data of the patient,

the method comprising:

1) training a neural network on numerous examples of (i) alleles data, corresponding (ii) diagnosed diseases, of a multiplicity of diseased patients so as to make a trained neural network that is fit, and that possesses a measure of goodness, to map (i) alleles data to (ii) diagnosed diseases; and

2) exercising the trained neural network on the alleles data of the particular individual patient to predict the susceptibility of the particular patient to at least one disease from among the diseases to which the neural network was trained.

13. A method of predicting at least one clinical result for a

particular individual patient in respect of alleles and/or SNP pattern data of the patient, the method comprising:

1) training a neural network on numerous examples of (i) alleles and/or SNP pattern data, and corresponding (ii) historical clinical results, for a multiplicity of patients so as to make a trained neural network that is fit, and that possesses a measure of goodness, to map (i) alleles and/or SNP pattern data to (ii) clinical results; and

2) exercising the trained neural network on the alleles and/or SNP pattern data of the particular individual patient to predict at least one clinical result for the particular patient from among the clinical results to which the neural network was trained.

14. The method according to claims 9, 10, 11, 12, or 13

wherein the training is automated by computerized programmed operations using a genetic algorithm.

15. The method according to claims 9, 10, 11, 12, or 13 wherein the training is automated by computerized programmed operations using a genetic algorithm reduced in computational complexity by including the steps of:

grouping alleles and/or characteristic SNP patterns into families as are defined by (i) having similar expression patterns, or (ii) being turned on and off by another gene, or (iii) both having similar expression patterns and being turned on and off by the same gene; and

starting training of the neural network with the genetic algorithm by using the families so created as single inputs to the neural network, the training with the genetic algorithm continuing repetitively until, families of greater and lesser significance being identified, it becomes computationally possible to train the neural network to genomic data consisting of individual alleles and/or characteristic SNP patterns;

wherein partitioning of all alleles and/or characteristic SNP patterns into families permits training of the neural network in a hierarchy of stages, first to the families and only then to the individual alleles and/or characteristic SNP patterns.

16. A method of training a neural network having a multiplicity M of inputs to extract information from genomic data having a great multiplicity of N variables, N >> M, unknown ones and unknown numbers of a majority of which N variables are both irrelevant and non-contributory to information that is extractable as desired output from a trained neural net,

the method thus being directed to training a neural network having only M inputs to extract information from N variables, N >> M, where, although many of the N variables are irrelevant or of much lesser relevance than others of the N variables, it is not known which, nor what number, of the N variables are so substantially irrelevant to extracting the information,

the method being of a general nature of an exercise of strategies of (i) divide and conquer while (ii) suppressing incorporation of substantially irrelevant variables until, finally, a neural network, nonetheless to having only M inputs, is trained to extract information from genomic data having a great multiplicity of N variables where M << N,

the method comprising:

organizing a great multiplicity of N genomic variables into M categories, called artificial genes, where M << N;

inputting a same set of N input values into each of these M categories as a functional block;

creating, by use of the M artificial genes and the N input values, (i) a vector of N values, or weights, for each of the M artificial genes, the weights being initially set randomly;

defining a dot (scalar) product of (i) the N-valued vector with (ii) an input vector of N genomic variables to create (iii) one single output value;

repeating the deriving of the dot product between successive (ii) input vectors each of a successive N genomic variables and (i) the vector of N values that are initially random, for each of the M functional blocks;

wherein this repeating of the deriving M times creates a filter vector, or artificial chromosome, of M values, which M values correspond to M genes in the artificial chromosome;

mapping, with a neural network, the created filter vector, or artificial chromosome, as an input vector so as to calculate a cost output value, the cost output value being a function of how similar the neural network output value is to a desired result, while also taking into consideration how many of the weights in the artificial genes are sufficiently below some predetermined threshold so as to be considered negligible;

optimizing the cost output value so as to create, by modifying the weights of each artificial gene, a particular artificial chromosome which, when fed as an input vector into the mapping neural net, causes the output values of said neural net to assume an optimal cost function;

wherein the number of inputs to the mapping neural net is decreased to M out of the N genomic variables, M << N;

wherein from the great multiplicity of N genomic variables, those variables which have greatest relevance to the optimal output of the mapping neural net are preferentially selected while those variables which have least relevance to the optimal output of the mapping neural network are preferentially discarded; and

wherein the great multiplicity of N genomic variables are divided into M categories, or artificial chromosomes, having similar functionality.

17. The method of training a neural network according to claim 16

wherein the optimizing of the vector inputs to the M functional blocks which have assigned to them a unique output value is by use of a genetic algorithm.

18. The method of training a neural network according to claim 16 directed to identifying

a statistically significant group of N genomic datums in the form of alleles and/or SNP patterns as these genomic datums affect given clinical results, which group is generally known as a clinically relevant alleles combination and/or characteristic SNP pattern as the case may be, from

genomic data of N variables.

19. A method of reducing the computational cost and complexity of the optimization of a neural network for application to a great multiplicity of N genomic datums by combining (i) preprocessing of N inputs into M outputs, (ii) feeding the M outputs as inputs into a more manageable neural network having only M inputs, with M << N, and (iii) training the neural network on the M inputs, the method comprising:

1) preprocessing a great multiplicity of N genomic datums into M functional blocks, called an artificial chromosome where each functional block is an artificial gene, suitably input to the neural network by steps of

a) constructing a plurality of artificial chromosomes each by choosing random numbers A_i of genomic datums suitably input to the neural network as artificial genes, $1 \le A_i \le N$, each such artificial gene thus consists of a group G_i of the original genomic datums,

b) repeating this process for each category i, $1 \le i \le M$,

c) assembling the union of these artificial genes as one of the plurality of the artificial chromosomes, each such chromosome thus consisting of some A variables grouped into M pieces G_i, $1 \le i \le M$, with $\Sigma A_i = A$, with each group G_i of genomic datums containing A_i variables,

d) training and exercising the neural network having M inputs on the M groups collectively comprising an artificial chromosome drawn from the plurality of artificial chromosomes, the M groups of the artificial chromosome collectively having A genomic datums, producing from this training and exercising one trial mapping;

e) performing the training and exercising in parallel for a number X times, once for each artificial chromosome constructed, each instance of training thus being performed for distinct groups of A genomic datums, thus producing X trial mappings, one for each of X artificial chromosomes;

f) determining for each of the X trial mappings an associated cost function; and

g) selecting, in consideration of the X cost functions, a one of the X trial mappings that is associated with one of the cost functions that is optimal; and

2) exercising the neural network a computationally tractable number X of times, M < X < N, on the great multiplicity of N genomic datums as are preprocessed into M inputs to the neural network.

20. The method according to claim 19 wherein at least the g) selecting is by application of a genetic algorithm.

21. A method of predicting drug interactions between two or more drugs for a given patient,

the method more particularly serving to predict an optimal drug dosage for a particular individual patient in respect of alleles and/or characteristic SNP pattern genomic data of the particular individual patient,

the method comprising:

1) training a neural network on numerous examples of (i) alleles and/or characteristic SNP pattern genomic data, and corresponding (ii) historical drug dosage results including optimal drug dosages, for a multiplicity of patients so as to make a trained neural network that is fit, and that possesses a measure of goodness, to map (i) alleles and/or characteristic SNP pattern genomic data to (ii) drug dosage results including optimal drug dosages, the training including steps of

(1a) producing an artificial chromosome by constructing such a filter with initial random values to pre-process the entire set of N genomic inputs into a filter of M inputs, M << N,

(1b) repeating the producing X times, where X is a computationally small number, to produce a set of X filters, (1c) using the set of X filters as input to a neural net which maps said signals to a desired clinical output,

(1d) determining a cost function from said mapping, and

(1e) using said cost function with a genetic algorithm to choose optimal filter values, and

(1f) optimizing the neural net for the fixed filter values obtained in (1e); and then

(2a) using the filter values corresponding to the first drug for the individual patient as inputs to a neural net which maps said signals to a desired clinical output for another drug;

(2b) optimizing this second neural net to produce the desired clinical output for the second drug with the input filter produced in (1e) held fixed;

(2c) using a standard numerical root finder to obtain a set of filter values which when used as inputs to the trained net obtained in (1f) produce a zero or near-zero output;

(2d) using said set of filter values produced in (2c) as inputs to the trained neural net obtained in (2b);

(2e) assembling two sets of filtered output signals as inputs to the trained neural net obtained in (2b), one from passing the given patient's genomic inputs through the filters obtained in (1e) the other by passing these same inputs through the filter(s) obtained from the root finding routine of (2c); and

(2f) identify as a measure of drug interaction the difference in the output of the neural net of (2b) using the input vectors as described in (2e).

22. A method of identifying a set of universal functional categories of genomic information, each universal functional category of genomic information being a set of genomic data that has a high probability of being relevant to more than one clinical variable of interest, the method comprising:

1) producing an artificial chromosome for one clinical variable of interest by

1a) constructing a filter with initial random values to pre-process the entire set of inputs to a single filtered signal,

1b) repeating the producing N times, where N is a computationally small number, to produce a set of N filtered signals,

1c) using the set of N filtered signals as input to a neural net which maps said signals to a desired clinical output,

1d) determining a cost function from said mapping; and

1e) using said cost function with a genetic algorithm to choose optimal filters; and then

2) repeating the 1) producing for Q clinical variables of interest, deriving Q optimal filters:

3) combining the Q optimal filters so produced via the steps of

3a) converting said Q filters obtained in (2) to binary filters by comparing each component of all filters to a predetermined threshold value, the component in question having value equal to 1 if the threshold is exceeded and zero otherwise,

3b) determining which of the binary filters are similar by performing the logical operation AND on pairs of filters,

3c) summing over the true values, and normalizing this sum in some manner, for example, the minimum of the either the first or second filter ANDed and summed with itself,

3d) joining filters by performing the logical operation OR upon them if the value produced in (3c) exceeds a predetermined threshold, and

3e) repeating the process described in (3c) and (3d) until no pair of filters has a threshold overlap, and

3f) identifying the resulting set of filters each of which filters is a universal functional category of genomic information, the set of filters being the set of universal functional categories of genomic information relevant to the more than one clinical variables of interest.

23. The method according to claim 22 further comprising:

4) refining each binary basis filter, the universal filter of interest, in the basis set to produce a non-binary basis filter set

having components consisting of probabilities that a gene which the component represents is actually a member of that basis filter set by steps of

4a) identifying for each of Q clinical variables of interest of step 1 that associated optimal filter obtained by step 2 that most completely overlaps the given binary basis filter in the basis set 3f, such overlap being determined by the mathematical sum of the bit-wise product of binary filter values,

4b) constructing N averages, each average being taken over Q values, each such value taken from the product of Q_1 and U_i, $1 \leq i \leq N$, with Q_i the ith component of the filter found in step 4a, and with U_1 the i^{tn} component of the universal filter of interest,

4c) identifying the corresponding collection of N clinical-variable-averaged binary filter/universal filter overlap values, which are the N averages found in step 4b, as a collection of probabilities that corresponding genomic data inputs are present in the closest binary universal filter, and

4d) identifying as a non-binary form of the universal filter those probabilities obtained in step 4c.

24. A method of using the universal functional categories of genomic information in accordance with claim 22 to predict the effect of a therapeutic regime, such as the administration of drugs, on a clinical output of interest, given the prior knowledge of the effect of said therapeutic regime on another, different clinical output, the method further comprising:

5) training a neural net to map these basis sets to the given therapeutic measure;

6) performing a root-finding technique to produce a representation of the patient's genome as affected by the desired therapeutic regime;

7) constructing a mapping neural network between a universal basis set of genomic inputs and a given clinical output of interest;

8) first feeding the corrected genomic inputs from step 6 performing through the network resulting from step 7 constructing,

and identifying a first network output as the predicted clinical output for the given patient as corrected for the desired therapeutic regime;

9) second feeding the patient's original genomic inputs, without application of the desired therapeutic regime, through the network resulting from step 7 constructing to produce a second network output; and

10) identifying the difference between the first network output obtained in step 8 and the second network output obtained in 9) as a measure of the effect of the desired therapeutic regime for the given patient.

25. The method of claim 24 exercised to predict the effect of each of two or more therapeutic regime(s) on a given clinical output.

26. A method of using the universal functional categories of genomic information in accordance with claim 24

wherein the inputs are genomic data such as specific alleles and/or characteristic SNP pattern(s),

wherein these inputs are used to produce an artificial chromosome, also called a filter,

wherein M filters are combined to produce a universal basis set of genomic inputs, and

wherein the universal basis set of genomic inputs is thus used to choose an optimal therapeutic regime for a given patient, wherein the method further comprises:

11) identifying potential problematic alleles and/or characteristic SNP pattern(s) known a priori;

12) constructing universal functional categories produced in step 3;

13) relating said universal functional categories to the problematic alleles and/or characteristic SNP pattern(s) by step 10; and

14) finding the effect of differing therapeutic regime by noting their effect upon these universal functional categories and

hence the effects of the problematic alleles and/or characteristic SNP pattern(s) by step 10.